Physica A 512 (2018) 57–67

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Impulse Weibull distribution for daily precipitation and climate change in China during 1961–2011

Choujun Zhan [a,b], Weiwen Cao [b], Junyu Fan [b], C.K. Tse [c,*]

[a] School of Computer, South China Normal University, China
[b] Nanfang College of Sun Yat-Sen University, Guangdong 510970, China
[c] Department of Electronic and Information Engineering, Hong Kong Polytechnic University, Hong Kong

HIGHLIGHTS

- A new Impulse Weibull (IWBL) probability distribution describing the daily precipitation is proposed.
- Statistic patterns of the daily surface precipitation data from the 3825 measurement sites from 1961 to 2011 in China is studied and follow IWBL distribution.
- Develop a method for generating the trends in the variation of the precipitation data.
- The trends in annual and daily precipitation over the period 1961 to 2011 is examined.

ARTICLE INFO

Article history:
Received 7 May 2018
Received in revised form 1 July 2018
Available online 30 July 2018

Keywords:
Precipitation
Climate change
Statistics property
Weibull distribution
Weather extreme

ABSTRACT

Based on a newly developed dataset containing daily precipitation in China at 0.5° intervals of longitude and latitude over the period 1961 to 2011, we find that statistical patterns of the daily surface precipitation data from the 3825 measurement sites can be described by a new Impulse Weibull (IWBL) probability distribution. By applying appropriate parameter identification techniques, we estimate all the parameters of the IWBL distribution for each site. We also examine the trends in annual and daily precipitation over the period 1961 to 2011. Results show that the probability of rainy days has decreased over time for over 90% the surface area of China, and that the extreme precipitation and annual precipitation have decreased for most of the area in China.

1. Introduction

Climate change, be it a scientific research or practically essential topic of study, has been pursued since ancient times. Studies of climate data, e.g., precipitation, temperature, snow falls, etc, have attracted increasing attention from the scientific community, agricultural industry, transport agencies, policy makers as well as other stakeholders. Vital to the study of climate is the collection of high quality measured data, for which a number of large-scale projects were launched aiming to archive accurate weather data and to construct grid dataset. In 1963, the World Meteorological Organization (WMO) established the World Weather Watch program to coordinate the observational capability of surface and satellite observations. In 1989, the Global Prediction Climatology Centre (GPCC) was established for developing reliable land surface dataset [1]. Five years later, in order to produce global analyses of area- and time-average precipitation, the Global

* Corresponding author.
E-mail addresses: zchoujun2@mail.com (C. Zhan), michael.tse@polyu.edu.hk (C.K. Tse).
URL: http://cktse.eie.polyu.edu.hk/ (C.K. Tse).

https://doi.org/10.1016/j.physa.2018.07.033
0378-4371/

Precipitation Climatology Project (GPCP) was established by the World Climate Researcher Programme [2]. Then, over the next decade (1994–2004), meteorologists from various academic institutions and organizations had coordinated the construction of monthly precipitation dataset, including the widely used monthly precipitation dataset, GPCP Versions 1 and 2, on $2.5^\circ \times 2.5^\circ$ latitude–longitude grids [3,4] and a monthly global (including land and ocean) precipitation dataset derived from over 17,000 stations collected in the Global Historical Climatology Network (GHCN) and the Climate Anomaly Monitoring System (CAMS) [5]. With the availability of advanced satellite precipitation estimation techniques, the first dedicated precipitation satellite was launched by the Tropical Rainfall Measuring Mission (TRMM) in 1997, and ten years later, the TRMM Multi-satellite Precipitation Analysis (TMPA) was developed and provided a dataset at fine scale covering the latitude band 50°N–50°S for the period from 1998 to 2007 [6]. However, monthly datasets are still deficient in many aspects for the analysis of climate change, such as in the length of growth seasons, the frequency of precipitation, and the frequency and duration of heat wave [7]. Then, a number of regional gridded daily precipitation datasets have been constructed by meteorologists from the USA, Europe, India, and Australia. One of the pioneering works, the One-Degree-Daily (1DD) technique for producing globally complete daily estimates of precipitation on $1^\circ \times 1^\circ$ grids from observed data (40°N–40°S), was implemented by Huffman and his collaborators [8]. However, this dataset only covers a short time from 1997 to 1999. A high-resolution gridded dataset (1950–2006) of daily climate over Europe (E-OBS) has been reported in 2008 [9], and three years later, a new dataset (ERA-Interim) has been presented by the European centre for Medium-Range Weather Forecasts (ECMWF) [10]. Daily precipitation dataset over the Indian region on 0.5° interval grids was also constructed from over 6000 rain-gauge data [11].

In China, despite being a country with highly populated regions and having a great need for monitoring climate data, relatively little effort has been made to construct gridded climate data, and the development of precipitation analyses over China has been slow compared with other countries. Until 2009, a $0.5^\circ \times 0.5^\circ$ grid daily temperature dataset for the period of 1961 to 2005 over the Chinese mainland based on the interpolation from only 751 observing stations in China has been developed [12]. In 2010, through the operation with the Climate Prediction Center affiliated to National Ocean and Atmospheric Administration (NOAA), a dataset with 0.5° grid resolution has been established based on the daily precipitation observations of over 2419 gauges [13]. Another dataset has been constructed and published by China Meteorological Administration (CMA; http://data.cma.cn/en) in 2013 [14]. This dataset (CN05.1) is based on the interpolation from over 2400 observing stations in China between 1961 and 2005. However, in order to generate useful projections without inheriting variations of long-term averages, it has been widely known that high-resolution data of a sufficiently long time period should be used, and a 50-year period is often the basic threshold. Thus, a high-resolution (at least $0.5^\circ \times 0.5^\circ$) daily precipitation dataset of China for a sufficiently long time period (at least 50 years) is still not available in a readily accessible form. In our study, we have constructed a dataset from 3825 stations covering the entire mainland area of China at a resolution of $0.5^\circ \times 0.5^\circ$ grid from 1961 to 2011.

The form of probability distribution function of daily precipitation is valuable for many purposes, e.g., in estimation of missing data, construction of a precipitation generator [15–17]. Simple distribution functions such as Gamma, Lognormal and Weibull functions have been used to describe weather data [18,19]. For instance, using Maximum Likelihood Estimation (MLE) [20,21], the parameters of the Weibull distribution can be readily estimated from a set of random numbers [22–24]. Owing to their ease of parameter estimation, simple distribution functions have been chosen by meteorologists for fitting weather data, despite the knowingly significant fitting errors. For daily precipitation data, we show that these simple distribution functions fail to describe the statistical characteristics well. When more complex distribution functions are used, the parameter estimation procedure is inevitably complicated and often involves a nonlinear optimization process with multiple local minima.

In this paper, we introduce a modified Weibull distribution, which we refer to as *Impulse Weibull distribution* (IWBL). We show that a wide range of precipitation characteristics can be described by specifying only three parameters, namely, shift parameter α, shape factor κ, and scale parameter λ. These three parameters are sufficient to specify the available precipitation and to enable assessment and evaluation of precipitation characteristics, such as average precipitation and extreme precipitation. A high-resolution dataset from the China Meteorological Administration is utilized in this study. This dataset consists of daily precipitation data observed from 3825 sites ($0.5^\circ \times 0.5^\circ$ grid points) covering the entire Chinese mainland area for the period 1961 to 2011. The parameters of the IWBL distribution are determined from the precipitation distribution statistics calculated from the data for a total of 50 years. We also develop a parameter estimation method to generate optimal parameter sets for all the 3825 sites. The daily precipitation data in relative frequency format calculated from the time-series and the IWBL-representative data are compared. The overall error of 3.79% shows that the IWBL-representative data are accurate representation of the time-series data. Furthermore, we introduce a method for studying extreme climates [25–27]. Our results show that the probability of the heavy and extreme daily precipitation events has increased over time in the mid-reach of the Yellow River, but has decreased in all other areas of China. The average annual precipitations from these calculated Impulse Weibull-representative parameters are also obtained. Results indicate that the average precipitation has decreased for over 90% of Chinese mainland area. We also apply the Theil–Sen estimator to project the probability of extreme precipitation and average precipitation in the future.

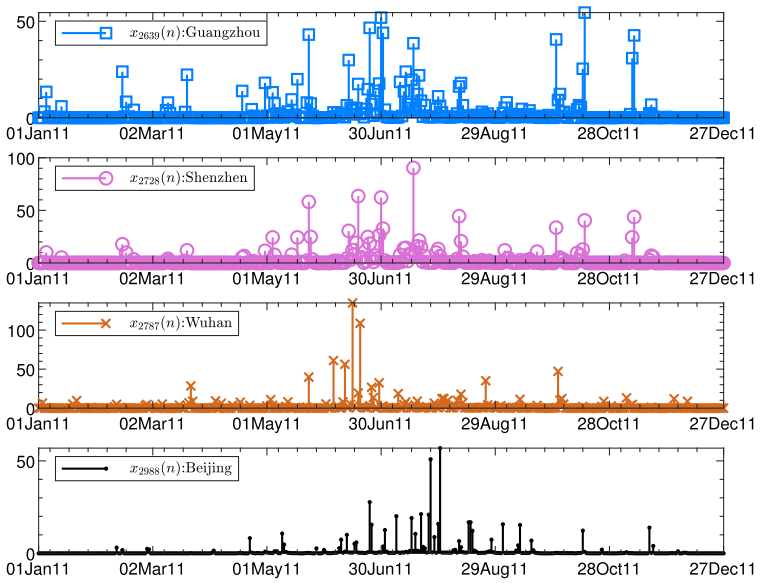


Fig. 1. Daily precipitation records from January 1, 2011 to December 31, 2011 of four major cites in China: Guangzhou (blue square), Shenzhen (orchid circle), Wuhan (brown cross) and Beijing (black point).

2. Preliminaries of the data and Impulse Weibull Distribution

2.1. Daily precipitation data

We have constructed a dataset of daily precipitation over the period 1961 to 2011 from 3825 measurement sites covering the entire mainland China. This is currently the largest dataset of daily precipitation in China with a space resolution of 0.5° intervals over a 30-year period. The measured data are available in time-series format, in which each data point represents a cumulative precipitation over some time period and the resolution of the series is 1 day. Then, there are 3825 different time series for the different locations and each time series contains 18 627 data points. Four data segments including daily precipitation record of four major cities are shown in Fig. 1. Moreover, in some other instances, precipitation data may instead be available in frequency distribution format, in which the frequency of the precipitation falls within various ranges (bins) are given. Here, we define the following variables:

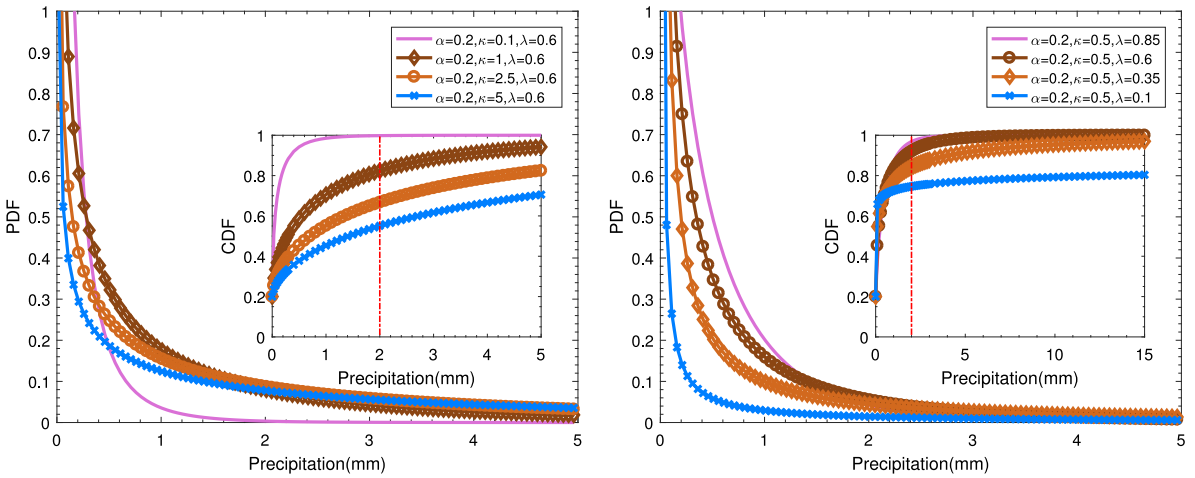
- x is the "true" daily precipitation;
- $x_i(n)$ $(1 \leq i \leq I$ and $1 \leq n \leq N)$ is the gridded daily precipitation of the ith area in the nth day, where $I = 3825$ and $N = 18\,627$.

Note that x and $x_i(n)$ have different physical meanings. Specifically, x represents the true value of daily precipitation, which is a real non-negative number, i.e., $x \in R^+$. However, $x_i(n)$ is the measured data which has been quantized. Here, the precision of gridded daily precipitation records is 0.1 mm, i.e., $x_i(n) = 0.1k$ mm, where $k \in Z^+$. For instance, on January 1, 1961, the actual daily precipitation of Guangzhou may be $x = 1.346$ mm. After being quantized, we get a measured data $x_{2636}(1) = 1.3$ mm. In a sunny day, the daily precipitation record is $x_i(n) = 0$ mm; otherwise $x_i(n) > 0$ mm. The daily precipitation can be considered as a random process. Ungrouped measured data are difficult to be read in any meaningful sense. Therefore, the precipitation data have to be grouped in order to show the important aspects of a distribution [28]. After grouping, the data in time-series format can be converted into frequency distribution format.

Fig. 1 shows the daily precipitation of four major Chinese cities (Guangzhou, Shenzhen, Wuhan and Beijing) from January 1 to December 31, 2011. From these data, we can derive the annual precipitation and standard deviation for a location. For instance, the annual precipitation and standard deviation of Wuhan are 1068 mm and 11.15 mm, respectively, and are 1124 mm and 8.02 mm for Guangzhou. Here, a large standard deviation with a high annual precipitation value indicates a large day-to-day variation in the precipitation, while a small standard deviation with a high precipitation value indicates small day-to-day variation but more rainy day spans.

2.2. Impulse Weibull Distribution

The Weibull (WBL) distribution, which is a generalization from exponential distribution, has generally been utilized for survival analysis [29]. It has been shown to give highly satisfactory fit to measured wind speed data [23,24,30–32]. The



(a) PDF of Impulse Weibull Distribution with varying κ

(b) PDF of Impulse Weibull Distribution with varying λ

Fig. 2. Examples of Impulse Weibull distribution with (a) $\alpha = 0.2$ and $\lambda = 0.6$, while κ increases from 0.1 to 5; (b) $\alpha = 0.2$ and $\kappa = 0.5$, while λ decreases from 0.85 to 0.1.

probability distribution function (PDF) of Weibull distribution is mathematically expressed by

$$p(\bar{x} < x < \bar{x} + dv) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(\frac{x}{\lambda})^{\kappa}}, \quad \text{for} \quad x \geq 0, \tag{1}$$

where x is the precipitation, $\bar{x}$ is a particular precipitation, and $p(\bar{x} < x < \bar{x}+dv)$ is the probability that the daily precipitation is between $\bar{x}$ and $\bar{x} + dv$. The cumulative distribution function is

$$P(x < \bar{x}) = 1 - e^{-(\frac{x}{\lambda})^{\kappa}}. \tag{2}$$

The complementary cumulative distribution function of Weibull distribution is a stretched exponential function. Based on the Weibull distribution, we propose the Impulse Weibull Distribution (IWBL) to describe the statistical properties of gridded daily precipitation data. The PDF of the IWBL distribution is

$$p(x) = \alpha\delta(x) + (1-\alpha)\frac{\kappa}{\lambda}\left(\frac{x}{\lambda}\right)^{\kappa-1} e^{-(\frac{x}{\lambda})^{\kappa}}\epsilon(x), \tag{3}$$

where $\delta(x)$ is the impulse function, and $\epsilon(x)$ is the step function. Here, x represents the daily precipitation record, and $x \leq 0$ means that the precipitation is 0. Note that there is an impulse at $x = 0$. From (3), we can derive the cumulative distribution function of the IWBL distribution as

$$P(x < \bar{x}) = \alpha + (1-\alpha)(1 - e^{-(\frac{x}{\lambda})^{\kappa}}), \quad x \geq 0. \tag{4}$$

where α, κ and λ are the parameters of the IWBL distribution. These parameters can be conveniently interpreted as follows:

- $\alpha \in (0, 1)$ is the *shift* parameter. $P(x = 0) = \alpha$ represents the probability of a sunny day, while $P(x > 0) = 1 - \alpha$ is the probability that the daily precipitation exceeds zero, i.e., rainy day;
- κ is the unitless *shape* parameter which causes the shape of the distribution to vary. By changing the value of κ, we can generate a set of curves with varying shapes to model precipitation distributions;
- $\lambda \in (0, \infty)$ is the *scale* parameter of the distribution, having the same unit as the precipitation.

We will show that the daily precipitation data x of the 3825 sites all follow the IWBL distribution. In applying (3) and (4), the probability can be interpreted as being relative (fractional or percentage) or absolute (number of data points). For example, $P(x > 0)$ in (4) can be interpreted as a fractional probability that daily precipitation exceeds zero or the number of days per year that the daily precipitation exceeds zero. The mean precipitation R_m is closely related to α, λ and κ, i.e.,

$$R_m = (1-\alpha)\lambda\Gamma\left(1 + \frac{1}{\kappa}\right) \tag{5}$$

where Γ is the usual gamma function. The shape factor κ is inversely related to the variance of precipitation (i.e., high κ values mean low variance, and vice versa). Fig. 2 shows how α, λ and κ affect the daily precipitation distribution. In particular, we make the following observations:

- A larger α means more sunny days.
- A smaller κ means less precipitation with larger amount, namely, fewer extreme events (heavy rain, rainstorm or downpour), as shown in Fig. 2(a). For instance, most measured precipitation data (99%) are below 2 mm (left side of the red dash line) with $\kappa = 0.1$ (black line). Moreover, for $\kappa = 5$, there are only about 40% of precipitation data below 1 mm (cyan cross line). The increase in κ also indicates an decrease in the average precipitation.
- A smaller λ means more extreme events, as shown in Fig. 2(b). For instance, most precipitation data (about 98%) are below 1 mm (left side of the blue dash line) with $\lambda = 0.85$. Moreover, for $\lambda = 0.1$, only about 60% precipitation data are below 1 mm and more than 20% precipitation data are above 15 mm. A larger λ also indicates a higher average precipitation.

In conclusion, larger λ and smaller κ correspond to fewer extreme events or more "fine weather" days, as can be seen directly from the CDF of IWBL distribution (4).

2.3. Parameter estimation of impulse Weibull Distribution

A common problem in model building is to identify the form of distribution that fits a given series of data. Here, given a specified distribution function which is known to provide satisfactory fit to the series of data, we find the values of all parameters of the distribution that give the smallest error, i.e., best fit. Here, we claim that the daily precipitation records of an area in China, such as in Guangzhou ($x_{2639}(n)$), follows the IWBL distribution. The parameters of IWBL distribution (α, λ and κ) are, however, to be estimated for each set of data. This is a very challenging task of model building, and can be described as an inverse problem in estimating the parameters from limited or sometimes only partially available experimental data [33–36].

For the normal Weibull distribution, only two parameters λ and κ are to be estimated. One commonly used method is to apply the least-squares fit to the observed distribution, as applied by Jamal to the wind speed data [22]. Furthermore, a modified maximum likelihood method has been attempted by Seguro and Lambert for parameter estimation [23], and a method using the gamma function to calculate the Weibull function parameters from the measured wind data in time-series format has been proposed by Celik [24]. However, these methods cannot be directly applied for estimating parameters in the IWBL distribution which differs significantly from the normal Weibull distribution. Here, we develop a method for estimating the parameters of the IWBL distribution from the daily precipitation dataset mentioned above.

The PDF and CDF of the IWBL distribution can be generated from (3) and (4) with a given set of parameters $\{\alpha, \lambda, \kappa\}$. Thus, the parameter estimation problem can be formulated as an optimization problem aiming to achieve good agreement (minimum error) between the output of the mathematical model and the frequency distribution format from the measured data [35]. The identification procedure is precisely an optimization problem:

$$\begin{aligned} &\mathrm{P}_0 : \min_{\theta} \int_{-\infty}^{\infty} \|p(x|\,\alpha, \lambda, \kappa) - p(x)\| dx, \\ &\text{s.t.} \quad \begin{cases} \text{(i)} & 1 \geq \alpha \geq 0, \\ \text{(ii)} & \kappa, \lambda \geq 0, \end{cases} \end{aligned} \tag{6}$$

where $\theta = \{\alpha, \lambda, \kappa\}$ is the set of parameters to be estimated, and $p(x)$ is the true probability distribution. Here, P_0 minimizes a cost function that measures the fitness of the model with respect to a given set of experimental (measured) data subject to a set of constraints. Also, $\|\cdot\|_l$ denotes the l-norm with $l > 0$. The optimal solution, if exists, is $\{\alpha^*, \lambda^*, \kappa^*\}$.

As the actual $P(x_i)$ and $p(x_i)$ cannot be obtained, we can only estimate the probability distribution from the measured data. Then, P_0 can be modified as

$$\begin{aligned} &\mathrm{P}_1 : \min_{\theta} \sum_{i=1}^{\infty} \left\| \int_{-\infty}^{0} p(x|\,\alpha, \lambda, k) dx - \alpha \right\| + \left\| \int_{x_i}^{x_{i+1}} p(x|\,\alpha, \lambda, \kappa) dx - P_r(x_i < x \leq x_{i+1}) \right\|, \\ &\text{s.t.} \begin{cases} \text{(i)} & 1 \geq \alpha \geq 0, \\ \text{(ii)} & \kappa, \lambda \geq 0. \end{cases} \end{aligned} \tag{7}$$

The application of this parameter estimation method requires that the precipitation data be in cumulative frequency distribution format. Thus, the time-series data must first be reformatted. Here, $n(x = x_i)$ presents the number of measured data, which is x_i. Then, we define the frequency distribution $p_i = P_r(x_i < x \leq x_{i+1}) = \frac{n(x=x_i)}{N}$, where N is the total number of time series data. From (4), we can get

$$\begin{aligned} &\mathrm{P}_2 : \min_{\theta} \sum_{i=1}^{\infty} \left((1-\alpha) \left(e\left(\frac{x_i}{\lambda}\right)^{\kappa} - e\left(\frac{x_{i+1}}{\lambda}\right)^{\kappa} \right) - p_i \right)^2 + (p_0 - \alpha)^2 \\ &\text{s.t.} \begin{cases} \text{(i)} & 1 \geq \alpha \geq 0, \\ \text{(ii)} & \kappa, \lambda \geq 0. \end{cases} \end{aligned} \tag{8}$$

where P_2 is a non-convex optimization problem. The RMS residual error can be used as a measure of "goodness-of-fit" of the distribution.

Evolutionary algorithms have been extensively used in parameter estimation. In this paper, a modified differential evolutionary algorithm [35] is proposed to solve the parameter estimation problem. The optimal values of IWBL parameters $\{\alpha_i^*, \kappa_i^*, \lambda_i^*\}$ are determined for the 3825 sites across China. Experimental results show that measured precipitation data in every site follows our IWBL distribution. Detailed results will be presented in Section 3.1.

2.4. Climate change analysis

Based on the IWBL distribution and the parameter estimation presented above, we develop a method for analyzing the climate change of an area (site). The procedure of the method is as follows:

1. Choose a time window with length N_w and a step n_s (N_w, $n_s \in Z^+$), where $N_w \leq N$ and $n_s \leq (N - N_w)$.
2. For site j, divide the time series of measured precipitation data $x_j(n)$ into $\left\lfloor \frac{N-N_w}{n_s} \right\rfloor$ boxes, $X_{j,k}$, i.e.,

$$X_{j,k} = \{x_j(n_s(k-1)+m) \,|\, m = 1, 2, \ldots, N_w\}, \tag{9}$$

   where $k = 1, 2, \ldots, M$ and $M = \left\lfloor \frac{N-N_w}{n_s} \right\rfloor$ is the largest integer below $\frac{N-N_w}{n_s}$.
3. Apply the parameter estimation method proposed in Section 2.3 to find the optimal parameters from the kth box by solving the following optimization problem:

$$\mathrm{P}_3 : \min_{\{\alpha_{j,k}, \lambda_{j,k}, \kappa_{j,k}\}} \left(p_{j,0} - \alpha_{j,k}\right)^2 + \sum_{i=1}^{\infty} \left((1-\alpha) \left(e\left(\frac{x_i}{\lambda}\right)^{\kappa} - e\left(\frac{x_{i+1}}{\lambda}\right)^{\kappa} \right) - p_{j,i} \right)^2$$
$$\text{s.t.} \begin{cases} \text{(i)} & 1 \geq \alpha_{j,k} \geq 0 \\ \text{(ii)} & \kappa_{j,k}, \lambda_{j,k} \geq 0 \end{cases} \tag{10}$$

   where $p_{j,i} = \frac{n(x=x_i)}{N_w}$ and $n(x = x_i)$ is the number of measured data with $x_i \in X_{j,k}$. By solving (10), we can find the optimal solution $\theta_{j,k}^* = \{\alpha_{j,k}^*, \kappa_{j,k}^*, \lambda_{j,k}^*\}$.
4. Derive the slope of the optimal parameters of each location. Define

$$\Phi = \begin{bmatrix} \left(\theta_{j,1}^*\right)' & \left(\theta_{j,1}^*\right)' & \cdots & \left(\theta_{j,1}^*\right)' \end{bmatrix}', \tag{11}$$

$$T = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ N_w & N_w + n_s & \cdots & N_w + M n_s \end{bmatrix}', \tag{12}$$

$$K = \begin{bmatrix} b_{j,\alpha} & b_{j,\lambda} & b_{j,\kappa} \\ k_{j,\alpha} & k_{j,\lambda} & k_{j,\kappa} \end{bmatrix}, \tag{13}$$

   Finally, obtain the change rate of each parameter by

$$K = \left(T'T\right)^{-1} T'\Phi. \tag{14}$$

   where $k_{j,\alpha}$, $k_{j,\lambda}$ and $k_{j,\kappa}$ are the change slope of the three key parameters, which can reflect the climate change.

We will apply this procedure to analyze the change of the three parameters $\{\alpha, \kappa, \lambda\}$ using the precipitation data of the 50 years, and will uncover the trends of extreme daily precipitation and annual precipitation in the following section.

3. Experimental results

3.1. Probability distribution of precipitation in China

The three-parameter IWBL function is fitted with the daily precipitation records from the 3285 different locations. The method proposed in Section 2.3 is applied to estimate the optimal IWBL parameters for each location. To assess the long-term effect, the precipitation data for the 50 years have been analyzed. The estimated shift, shape and scale parameters of the ith site are α_i, κ_i and λ_i, respectively. The IWBL probability density function generated using the estimated parameter values is compared with the probability density function generated from the data. Results are shown in Fig. 3, where the solid lines represent the PDF and CDF generated by the IWBL model, and the dash lines represent the CDF and PDF from the measured data. For instance, the optimal parameters of Guangzhou are $\alpha_{2639}^* = 0.3433$, $\lambda_{2639}^* = 0.1673$, and $\kappa_{2639}^* = 3.0427$, and those for Beijing are $\alpha_{2988}^* = 0.4094$, $\lambda_{2988}^* = 0.3382$, and $\kappa_{2988} = 0.4770$. Statistical results show that the mean value and standard deviation of the error between the estimated PDF and PDF obtained from the measured data is 0.0043 ± 0.0024, while the mean value and standard deviation of the error between the estimated CDF and the true CDF is 0.0040 ± 0.0025. Clearly, the distribution generated by IWBL agrees well with the measured probability density function for all cases. Moreover, when standard distribution functions are used, no satisfactory fit can be obtained for the daily precipitation data.

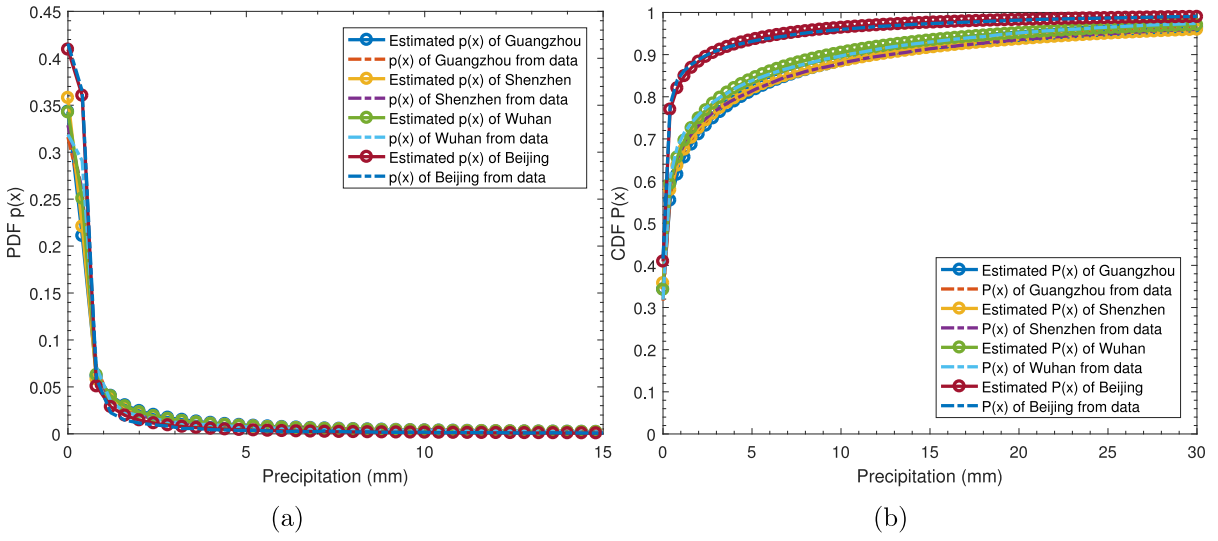
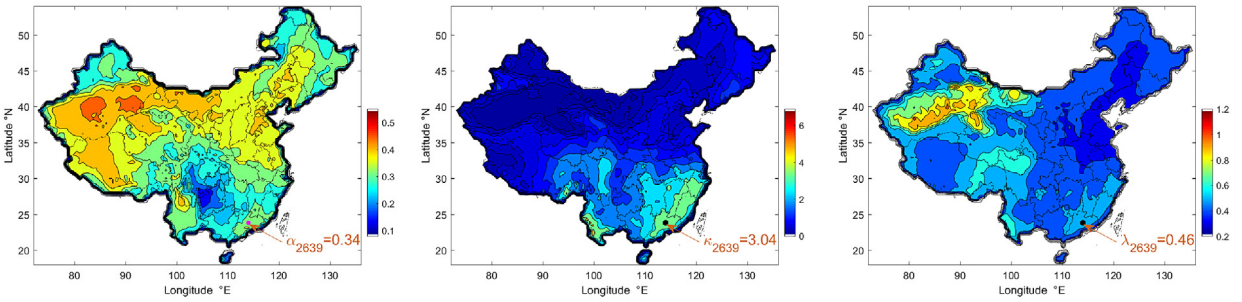


Fig. 3. IWBL fitting for data of four cities (Guangzhou, Shenzhen, Wuhan and Beijing). (a) Estimated PDF (solid line) and PDF from gridded precipitation data (dash line); (b) estimated CDF (solid line) and CDF from gridded precipitation data (dash line).



(a) 3825 50-year-long shift parameters α_i^* (b) 3825 50-year-long shape parameters κ_i^* (c) 3825 50-year-long scale parameters λ_i^*

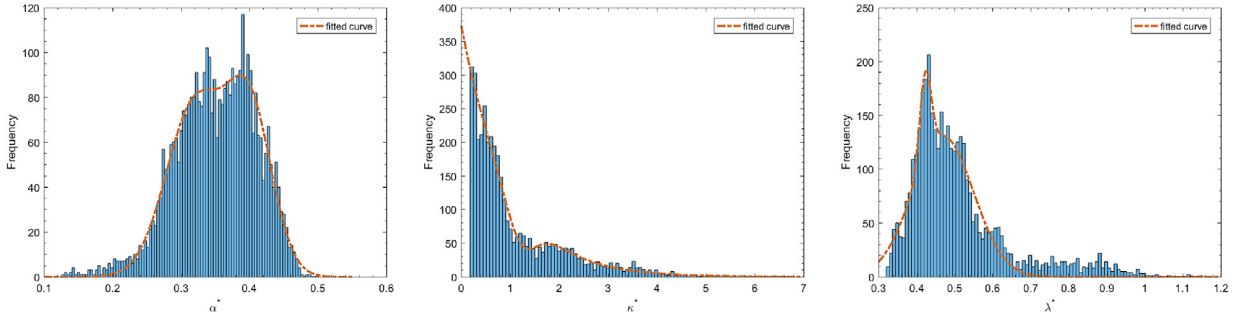


Fig. 4. (a)–(c) Contour maps of the three optimal parameter sets α_i^*, κ_i^* and λ_i^* with different colors representing different values; (d)–(f) histogram maps of the three parameters; red dot curves are curves fitted by Gaussian 2 function. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

We apply the optimization procedure to generate the optimal parameter values for the 3825 sites, and plot the contour maps of α^*, λ^* and κ^*, as given in Fig. 4(a)–(c). For instance, at the location of Guangzhou (113°E and 23°N), the value of α_{2639}^* is 0.3433. Then, we use a blue point to represent α_{2639}^* at 113°E and 23°N (Fig. 4(a)). Similarly, we plot the contour maps and the histograms of the three parameters, as given in Fig. 4(a)–(c) and (d)–(f), respectively. The red line is the fitted curve

generated by Gaussian 2 function, which is a multimodal distribution [37]. Results show that the values of shift parameter α lie between 0.25 to 0.4. The values of shape parameter κ are between 0 and 3. For κ larger than 3, the relative frequency distribution is more peaked and indicates a higher probability of having "fine" days. The values of scale parameter λ are mostly between 0 and 1. With constant α and κ, from Eq. (4), a small value of λ indicates that the probability of a heavy rainy day is high. When λ is close to 0, the cumulative frequency distribution is relatively flat. For instance, as shown in the inset of Fig. 2(b), when $\lambda = 0.1$, about 20% of the precipitation data are higher than 15 mm, while for $\lambda = 0.85$, only less than 4% of the precipitation data are higher than 15 mm. Thus, a smaller λ indicates a highly fluctuating precipitation pattern, whereas a larger λ indicates more regular and steady daily precipitation. Parameter α is the probability of a day being rainy, and hence effectively represents the number of rainy days in a year. Thus, Fig. 4(a) shows the distribution of the number of rainy days per year in China over the past 50 years. As can be seen from Fig. 4(a), Xinjiang Province has the smallest number of rainfall days per year, namely, more than 180 days of precipitation being less than 0.1 mm each year in the past 50 years. Areas along the border between Guizhou and Sichuan Provinces has the most rainy days, namely, more than 80% of the days per year having precipitation records higher than 0.1 mm. The values of parameters κ and λ influence the probability of occurrence of extreme rainfall events. The regions with larger κ and smaller λ have a relatively higher probability of occurrence of extreme rainfall weather. As can be seen from Fig. 4(b) and (c), the probability of occurrence of extreme rainfall weather is relatively high in the southwest of Guangdong Province, Fujian Province, Zhejiang Province, and Yunnan Province, while the probability of occurrence of extreme rainfall weather in the entire northern China is relatively low.

3.2. Climate change in China

Climate change has been studied in the form of frequent extreme climate events (e.g., heatwaves, floods, droughts and wildfires) which have profound impacts on our environment. By applying the method introduced in Section 2.3, we analyze the climate change in China. The time window used should be long enough to reflect statistical characteristics. Previous research has shown that climate experienced different decadal variation features. Hence, we select half a decade, i.e., $N_w = 1825$, as the duration of time window for our study. The time step should be chosen short enough to guarantee accuracy, but not excessively short to affect computational efficiency. Here, the time step is settled as 1.5 year, i.e., $n_s = 548$. Furthermore, the 50 years are divided into 30 segments: [1 1825]d, [548 2373]d, $\cdots$, [16440 18265]d, where 'd' represents day. Some examples are given in Fig. 5(a)–(c), which show the optimal parameter set $\{\alpha^*_{j,k}, \kappa^*_{j,k}, \lambda^*_{j,k}\}$ for the precipitation data measured at the 2639th site ($j = 2639$) in China. We first estimate the parameter of the 30 segments, namely, $\{\alpha^*_{2639,1}, \lambda^*_{2639,1}, \kappa^*_{2639,1}\}$, $\{\alpha^*_{2639,2}, \lambda^*_{2639,2}, \kappa^*_{2639,2}\}, \ldots, \{\alpha^*_{2639,30}, \lambda^*_{2639,30}, \kappa^*_{2639,30}\}$. Then, we utilize a linear fit (red line) to fit the parameters. Finally, we derive the gradients $k_{2639,\alpha} = 6.45 \times 10^{-5}$, $k_{2639,\lambda} = 0.0031$ and $k_{2639,\kappa} = -0.0017$. It can be seen that the shift parameter α for this site varies from 0.375 to 0.405 with an average of 0.39. It is interesting to note from Fig. 5(a)–(c) that the shape parameter κ has an increasing trend, particularly from year 1973 onward, while the scale parameter λ has a decreasing trend. The lowest value of the scale parameter is 0.49 which occurred in year 1995, whilst the highest value is 0.72 which occurred in year 1965. The scale parameter shows a decreasing trend throughout the 50 years.

To assess the long-term effect, the gradients of the 3825 sites are derived from the precipitation data by applying the method introduced in Section 2.4. In order to observe the gradients of the three IWBL parameters graphically, the long-term gradients for the 3825 sites for the 50 years are shown in the set of contour maps given in Fig. 5(d)–(f), and also in the histograms shown in Fig. 5(g)–(i). The red line is the fitted curve generated by Gaussian 2 function.

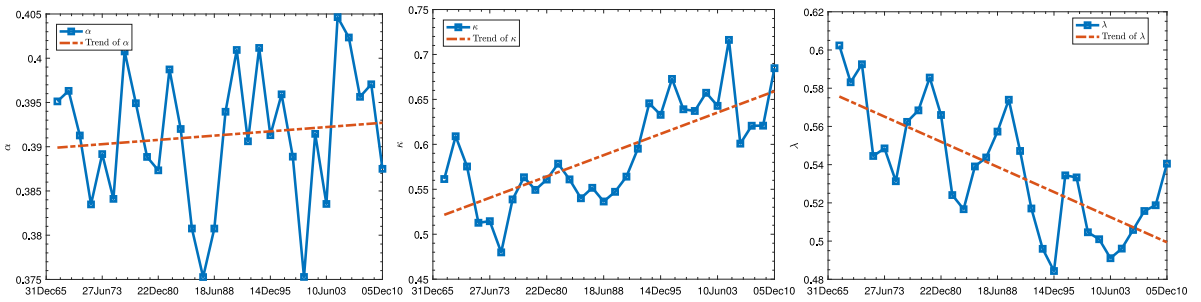
Note that the yellow region covers most developed areas in China. It can be observed that for much of China, the shift parameter α has an increasing trend over time (yellow region in Fig. 5(d)), which means that the number of sunny days increases (also shown in Fig. 5(g)). Only in a small part of Northwest China, α has a decreasing trend implying a higher occurrence of rainy days. Fig. 5(f) shows the contour map of $k_{j,\kappa}$, while Fig. 5(h) shows the corresponding histogram. The value of k_κ for about half of China increases, implying the number of extreme precipitation days increases. The contour maps and histograms of k_α are shown in Fig. 5(g) and (i), respectively. There is almost 50% of sites which have a positive k_α, i.e., falling within $[0 \quad 6 \times 10^{-5}]$. The other 50% of sites have a positive k_α, i.e., falling within $[-2 \times 10^{-5} \quad 0]$.

3.3. Climate change of extreme climate and average precipitation

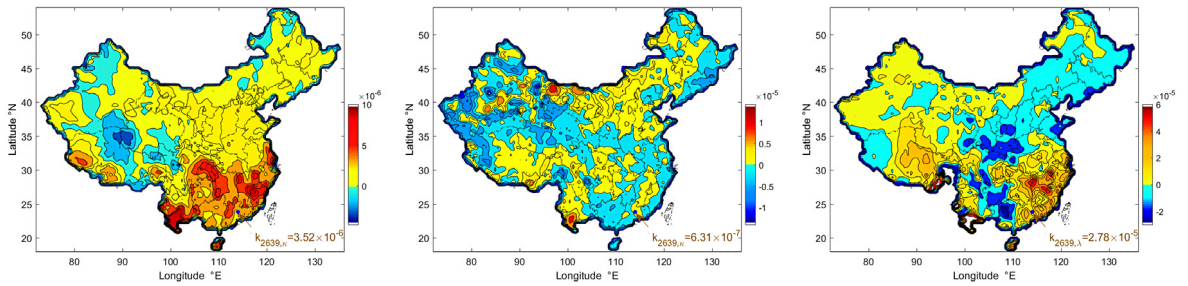
For ease of reference, and consistent with the terminology used in China, the level of precipitation can be roughly classified as follows:

- Light rain: precipitation between 0.1 and 9.9 mm in 24 h
- Rain: precipitation between 10 and 24.9 mm in 24 h
- Heavy rain: precipitation between 25.0 and 49.9 mm in 24 h
- Rainstorm: precipitation between 50.0 and 99.9 mm in 24 h
- Downpour: precipitation between 100.0 and 249.9 mm in 24 h
- Extraordinary storm: precipitation larger than 250.0 mm in 24 h

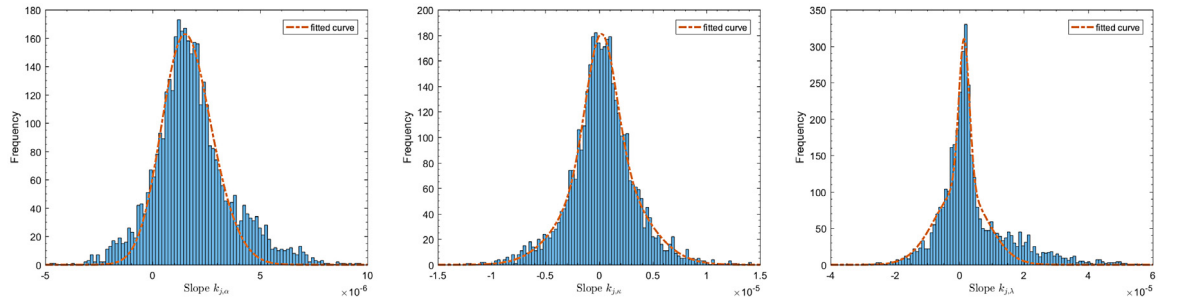
For simplicity, we consider "heavy rain", "rainstorm", "downpour" and "extraordinary storm" as extreme climate, for which the daily precipitation is larger than 30 mm.



(a) Slope $k_{2639,\alpha} = 6.45 \times 10^{-5}$ (b) Slope $k_{2639,\kappa} = 0.00316$ (c) Slope $k_{2639,\lambda} = -0.00175$



(d) 3825 slopes $(k_{j,\alpha})$ of shift parameters (e) 3825 slopes $(k_{j,\kappa})$ of shape parameters (f) 3825 slopes $(k_{j,\lambda})$ of scale parameters

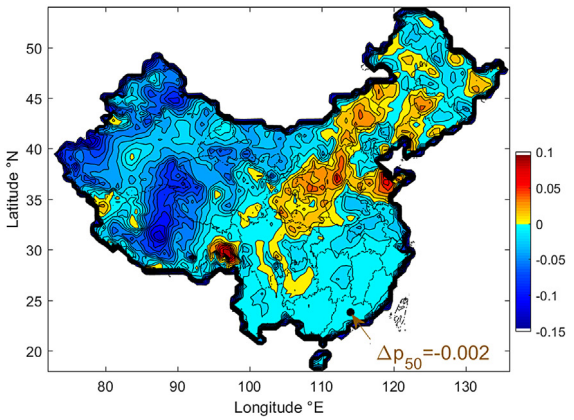


(g) (h) (i)

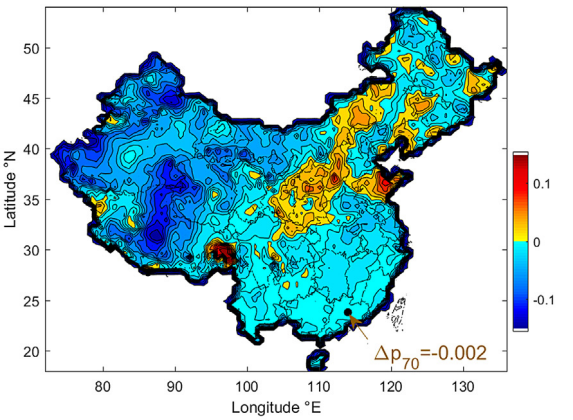
Fig. 5. (a)–(c) Estimated parameters in different time windows for the 2639th site; (d)–(f) gradients of the three parameters; (g)–(i) histograms of the gradients of the three parameters. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

In this section, the changes of the probabilities of extreme precipitation climate and average precipitation are analyzed. In China, "Heavy rain", "Rainstorm", "Downpour" and "Extraordinary storm" are treated as extreme weather. Hence, we set 30 mm as a threshold. In other words, $x \geq 30$ mm represents an extreme weather day. Using the IWBL parameters found, we derive the probability of daily precipitation being greater than 30 mm, i.e., $P(x \geq 30)$. First, for each of the 3825 locations, we calculate probability P_1 of the precipitation being greater than 30 mm in a particular year in each site. Then, we derive the probability P_2 of the precipitation being greater than 30 mm in a later year. The change of the probability of extreme precipitation is defined as $\Delta P = P_2 - P_1$, which is shown in Fig. 6(a) for years 1961 and 2011. We can then estimate the three IWBL parameters in a future year by Theil–Sen estimator [38]. In particular, we can derive the probability of extreme precipitation climate P_3 and the difference between P_3 and P_1. After performing extensive calculation, results show that in most areas of China, the probability of extreme precipitation decreases in the last 50 years and will decrease in the future, except in the mid-reach of the Yellow River to Inner Mongolia.

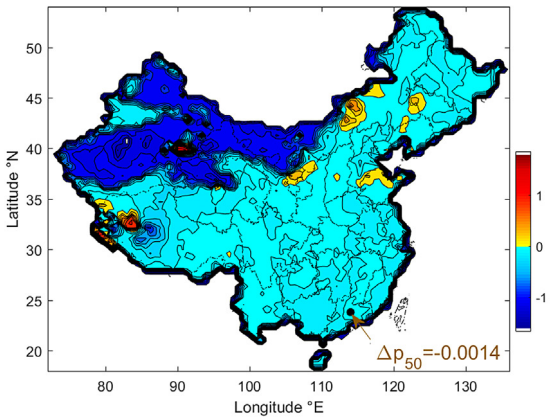
Using the estimated parameters, we can also derive the average precipitation from Eq. (5). We define Rm_1, Rm_2 and Rm_3 as the average precipitation in years 1961, 2011, and 2031, receptively. Fig. 6(c) and (d) show the contour maps of $(Rm_2 - Rm_1)$ and $(Rm_3 - Rm_1)$, respectively. Results shown that average precipitation decreases for more than 90% of the Chinese mainland area. The meteorologist always use linear regression to analyze the annual average precipitation and the frequency of heavy rain. For instance, the regression line of extreme precipitation days is represented as $x(t) = kt + b$, where



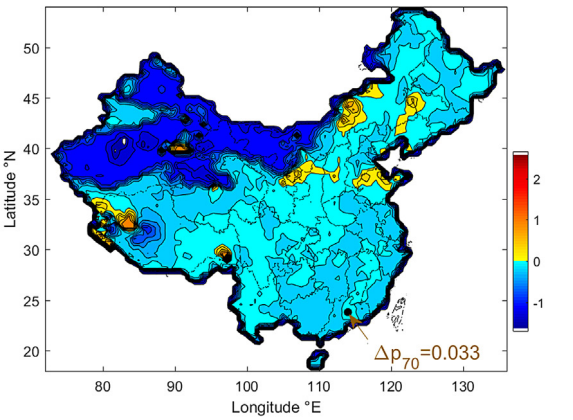
(a) Change of probability of heavy precipitation in 50 years



(b) Change of probability of heavy precipitation in 70 years



(c) Change of average precipitation in 50 years



(d) Change of average precipitation in 70 years

Fig. 6. (a) Change of probability of extreme precipitation climate between 1961 and 2011; (b) change of probability of extreme precipitation climate between 1961 and 2031; (c) average precipitation change between 1961 and 2011; (d) average precipitation change between 1961 and 2031.

$x(t)$ is number of annual extreme precipitation days and k is the slope of regression line. If the estimated slope k is positive, the trend will increase forever and as time elapse, the number of annual extreme precipitation day will larger than 365 days, which is impossible. However, our model is a probability model controlled by three parameters α, κ and λ. We derive the probability of extremely from Impulse Weibull Distribution, which cannot larger 1 (equal to 365 days) and is more reliable. Similar advantages can be found in analyzing annual average precipitation and other properties.

4. Conclusion

The precipitation data adopted for the present study has been constructed from 3825 measurement sites covering almost the entire China. These sites have retained records of the daily precipitation observations since 1961. It is believed that weather data for not less than 30 years should be adopted as shorter periods, less than 30 years, may inherit variations from the long-term average. Thus, record of longer periods can yield quantitative, representative as well as persuasive results. Numerical estimations using the Impulse Weibull Distribution (IWBL) three-parameter function to describe the daily precipitation frequency distribution for a given set of dataset over the period of 50 years (1961–2011) has been carried out. The values of the parameters for 3825 different locations are presented and examined. Based on the IWBL model, we have analyzed the climate change in China. We are also interested in the statistical properties of the heavy precipitation cases as they are important for modeling and understanding the weather dynamics. Results show that the probability of extreme climate decreases in most area of China. Furthermore, the average precipitation of over 90% of Chinese mainland has decreased in the last 50 years and will continue to decrease in the future. We have also applied the proposed method to other daily precipitation data such as those of the United States. Initial results show that the IWBL function fits well. Furthermore,

deep learning methods are known to offer advantages in tackling certain problems, such as classification, regression, and feature learning [39,40]. Several deep learning networks can be utilized for time-series prediction and regression, e.g., long short-term memory (LSTM) deep learning networks and Generative Adversarial (GAN) Networks. Future work may involve application of deep learning methods in predicting and analyzing climate changes.

Acknowledgments

This work was supported by National Science Foundation of China (61703355) and Hong Kong Polytechnic University, Hong Kong under Grants G-YBKB and GYBAT.

References

[1] U. Schneider, A. Becker, P. Finger, A. Meyer-Christoffer, M. Ziese, B. Rudolf, GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle, Theor. Appl. Climatol. 115 (1–2) (2014) 15–40.
[2] P.A. Arkin, P. Xie, The global precipitation climatology project: first algorithm intercomparison project, Bull. Am. Meteorol. Soc. 75 (3) (1994) 401–419.
[3] P. Xie, B. Rudolf, U. Schneider, P.A. Arkin, Gauge-based monthly analysis of global land precipitation from 1971 to 1994, J. Geophys. Res.: Atmos. 101 (D14) (1996) 19023–19034.
[4] R.F. Adler, G.J. Huffman, A. Chang, R. Ferraro, P.-P. Xie, J. Janowiak, B. Rudolf, U. Schneider, S. Curtis, D. Bolvin, et al., The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979–present), J. Hydrometeorol. 4 (6) (2003) 1147–1167.
[5] M. Chen, P. Xie, J.E. Janowiak, P.A. Arkin, Global land precipitation: A 50-yr monthly analysis based on gauge observations, J. Hydrometeorol. 3 (3) (2002) 249–266.
[6] G.J. Huffman, D.T. Bolvin, E.J. Nelkin, D.B. Wolff, R.F. Adler, G. Gu, Y. Hong, K.P. Bowman, E.F. Stocker, The trmm multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales, J. Hydrometeorol. 8 (1) (2007) 38–55.
[7] M.J. Menne, I. Durre, R.S. Vose, B.E. Gleason, T.G. Houston, An overview of the global historical climatology network-daily database, J. Atmos. Ocean. Technol. 29 (7) (2012) 897–910.
[8] G.J. Huffman, R.F. Adler, M.M. Morrissey, D.T. Bolvin, S. Curtis, R. Joyce, B. McGavock, J. Susskind, Global precipitation at one-degree daily resolution from multisatellite observations, J. Hydrometeorol. 2 (1) (2001) 36–50.
[9] N. Hofstra, M.R. Haylock, A. Tank, E. Klok, P. Jones, M. New, A european daily high-resolution gridded data set of surface temperature and precipitation, J. Geophys. Res. D Atmos. 113 (2008) 20.
[10] D.P. Dee, S. Uppala, A. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. Balmaseda, G. Balsamo, P. Bauer, et al., The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Q. J. R. Meteorol. Soc. 137 (656) (2011) 553–597.
[11] M. Rajeevan, J. Bhate, A high resolution daily gridded rainfall dataset (1971–2005) for mesoscale meteorological studies, Current Sci. (2009) 558–562.
[12] Y. Xu, X. Gao, Y. Shen, C. Xu, Y. Shi, F. Giorgi, A daily temperature dataset over China and its application in validating a RCM simulation, Adv. Atmos. Sci. 26 (4) (2009) 763–772.
[13] Y. Shen, M. Feng, H. Zhang, F. Gao, Interpolation methods of China daily precipitation data, J. Appl Meteorol. Sci. 3 (2010) 005.
[14] X.-J. Gao, A gridded daily observation dataset over China region and comparison with the other datasets, Diqiu Wuli Xuebao 56 (4) (2013) 1102–1111.
[15] M.O. Grondona, G.P. Podestá, M. Bidegain, M. Marino, H. Hordij, A stochastic precipitation generator conditioned on ENSO phase: a case study in southeastern South America, J. Clim. 13 (16) (2000) 2973–2986.
[16] C. Kilsby, P. Jones, A. Burton, A. Ford, H. Fowler, C. Harpham, P. James, A. Smith, R. Wilby, A daily weather generator for use in climate change studies, Environ. Model. Softw. 22 (12) (2007) 1705–1719.
[17] A. Baxevani, J. Lennartsson, A spatiotemporal precipitation generator based on a censored latent gaussian field, Water Resour. Res. 51 (6) (2015) 4338–4358.
[18] S. Michaelides, F. Tymvios, T. Michaelidou, Spatial and temporal characteristics of the annual rainfall frequency distribution in cyprus, Atmos. Res. 94 (4) (2009) 606–615.
[19] D. Stephenson, K.R. Kumar, F. Doblas-Reyes, J. Royer, F. Chauvin, S. Pezzulli, Extreme daily rainfall events and their impact on ensemble forecasts of the Indian monsoon, Mon. Weather Rev. 127 (9) (1999) 1954–1966.
[20] D.R. Cox, O. Barndorff-Nielsen, Inference and Asymptotics, vol. 52, CRC Press, 1994.
[21] L. Wasserman, All of Statistics: a Concise Course in Statistical Inference, Springer Science & Business Media, 2013.
[22] M. Jamil, Wind power statistics and evaluation of wind energy density, Wind Eng. (1994) 227–240.
[23] J. Seguro, T. Lambert, Modern estimation of the parameters of the Weibull wind speed distribution for wind energy analysis, J. Wind Eng. Ind. Aerodyn. 85 (1) (2000) 75–84.
[24] A.N. Celik, Energy output estimation for small-scale wind power generators using Weibull-representative wind data, J. Wind Eng. Ind. Aerodyn. 91 (5) (2003) 693–707.
[25] A. Rakshit, Analysis of Time Series Climate Data Using the Tools of Wavelet Transform, Random Matrix Theory and Multifractal Detrended Fluctuation Analysis (Ph.D. thesis), Indian Institute of Science Education and Research Kolkata, 2014.
[26] R.O. Weber, P. Talkner, Spectra and correlations of climate data from days to decades, J. Geophys. Res.: Atmos. 106 (D17) (2001) 20131–20144.
[27] M. Mudelsee, M. Börngen, G. Tetzlaff, U. Grünewald, No upward trends in the occurrence of extreme floods in central Europe, Nature 425 (6954) (2003) 166–169.
[28] P.J. Bickel, K.A. Doksum, Mathematical Statistics: Basic Ideas and Selected Topics, vol. 2, CRC Press, 2015.
[29] W. Weibull, Wide applicability, J. Appl. Mech. 103 (730) (1951) 293–297.
[30] C. Justus, W. Hargraves, A. Yalcin, Nationwide assessment of potential output from wind-powered generators, J. Appl. Meteorol. 15 (7) (1976) 673–678.
[31] I.Y. Lun, J.C. Lam, A study of Weibull parameters using long-term wind observations, Renew. energy 20 (2) (2000) 145–153.
[32] P. Ramírez, J.A. Carta, Influence of the data sampling interval in the estimation of the parameters of the Weibull wind speed probability density distribution: a case study, Energy Convers. Manage. 46 (15) (2005) 2419–2438.
[33] C. Zhan, B.Y.S. Li, L.F. Yeung, Structural and practical identifiability analysis of S-system, IET Syst. Biol. 9 (6) (2015) 285–293.
[34] C. Zhan, L.F. Yeung, Parameter estimation in systems biology models using spline approximation, BMC Syst. Biol. 5 (1) (2011) 14.
[35] C. Zhan, W. Situ, L.F. Yeung, P.W.M. Tsang, G. Yang, A parameter estimation method for biological systems modelled by ode/dde models using spline approximation and differential evolution algorithm, IEEE/ACM Trans. Comput. Biol. Bioinform. (TCBB) 11 (6) (2014) 1066–1076.
[36] K.-H. Cho, S.-M. Choo, S. Jung, J. Kim, H.-S. Choi, J. Kim, Reverse engineering of gene regulatory networks, IET Syst. Biol. 1 (3) (2007) 149–163.
[37] L. Cobb, P. Koppstein, N.H. Chen, Estimation and moment recursion relations for multimodal distributions of the exponential family, J. Amer. Statist. Assoc. 78 (381) (1983) 124–130.
[38] R. Fernandes, S.G. Leblanc, Parametric (modified least squares) and non-parametric (Theil–Sen) linear regressions for predicting biophysical parameters in the presence of measurement errors, Remote Sens. Environ. 95 (3) (2005) 303–316.
[39] H. Zhang, J. Li, Y. Ji, H. Yue, Understanding subtitles by character-level sequence-to-sequence learning, IEEE Trans. Ind Inform. 13 (2) (2017) 616–624.
[40] H. Zhang, X. Cao, J.K. Ho, T.W. Chow, Object-level video advertising: an optimization framework, IEEE Trans. Ind. Inform. 13 (2) (2017) 520–531.